The contribution of object size, manipulability, and stability on neural responses to inanimate objects

Caterina Magri[a,1*], Talia Konkle[b] and Alfonso Caramazza[b,c]

[a] Department of Psychology, Harvard University, Cambridge, MA 02138, USA

[b] Department of Psychology, Harvard University, Cambridge, MA 02138, USA

[c] Center for Mind/Brain Sciences (CIMeC), University of Trento, 38068 Rovereto (TN), Italy

[1]Department of Cognitive Science, Johns Hopkins University, Baltimore, MD 21218

[*]*Corresponding author.* Department of Cognitive Science, 247 Krieger Hall, 3400 N. Charles Street, Baltimore, MD 21218, USA. Email address: cmagri1@jhu.edu.

Abstract

In human occipitotemporal cortex, brain responses to depicted inanimate objects have a large-scale organization by real-world object size. Critically, the size of objects in the world is systematically related to behaviorally-relevant properties: small objects are often grasped and manipulated (e.g., forks), while large objects tend to be less motor-relevant (e.g., tables), though this relationship does not always have to be true (e.g., picture frames and wheelbarrows). To determine how these two dimensions interact, we measured brain activity with functional magnetic resonance imaging while participants viewed a stimulus set of small and large objects with either low or high motor-relevance. The results revealed that the size organization was evident for objects with both low and high motor-relevance; further, a motor-relevance map was also evident across both large and small objects. Targeted contrasts revealed that typical combinations (small motor-relevant vs. large non-motor-relevant) yielded more robust topographies than the atypical covariance contrast (small non-motor-relevant vs. large motor-relevant). In subsequent exploratory analyses, a factor analysis revealed that the construct of motor-relevance was better explained by two underlying factors: one more related to manipulability, and the other to whether an object moves or is stable. The factor related to manipulability better explained responses in lateral small-object preferring regions, while the factor related to object stability (lack of movement) better explained responses in ventromedial large-object preferring regions. Taken together, these results reveal that the structure of neural responses to objects of different sizes further reflect behavior-relevant properties of manipulability and stability, and contribute to a deeper understanding of some of the factors that help the large-scale organization of object representation in high-level visual cortex.

Highlights

- Examined the relationship between real-world size and motor-relevant properties in the structure of responses to inanimate objects.
- Large scale topography was more robust for contrast that followed natural covariance of small motor-relevant vs. large non-motor-relevant, over contrast that went against natural covariance.
- Factor analysis revealed that manipulability and stability were, respectively, better explanatory predictors of responses in small- and large-object regions.

Keywords

object representation; occipitotemporal cortex; functional magnetic resonance imaging, real-world size

1. Introduction

To recognize objects in the real world, large portions of the brain are utilized, including the ventral visual stream (Goodale and Milner, 1992; Ishai et al., 1999). Within this cortex there is a highly consistent organization of neural response tuning, with a tripartite large-scale topography distinguishing responses to animate entities, large inanimate objects, and small inanimate objects (Chao et al., 1999; Konkle and Caramazza, 2013; Proklova et al., 2016; Julian et al., 2017; Grill-Spector and Weiner, 2014). Some proposals have raised the possibility that the large-scale divisions in this cortex arise to support different behavioral needs (e.g., manipulating objects vs navigating in environments), linked to distinct underlying long-range brain networks (e.g., Mahon and Caramazza, 2011; Konkle and Caramazza, 2017). Thus, in the present work we examined how both the size of an object and its motor-relevance can explain brain responses to object categories.

Intuitively, the relationship between an object's size and the degree of interaction it affords in the world is evident: small objects tend to be graspable and easy to hold up, while larger objects tend to be used as support or for navigation. However, other relationships are possible, including not usually manipulated small objects like picture frames, and interactive large objects like pianos. Here, we operationalize the dimension dealing with object interaction as *motor-relevance*. While similar conceptually to the property of manipulability that has been investigated in the past (Kellenbach et al., 2003; Boronat et al., 2003; Mahon et al., 2007; Campanella et al., 2010; Kalénine and Buxbaum, 2016), motor-relevance spans a wider definition in that it includes actions performed not only with the hands, but with other parts of the body as well. With this broader conceptualization, size can more easily dissociate from degree of interaction: large objects like swing sets and wheelbarrows tend to involve movements beyond hand-performed actions, while for small objects such as picture frames and smoke alarms there is no clear motor interaction associated.

To date, there is substantial evidence that motor-relevance is an important construct for visual object responses, particularly along the lateral aspect of the occipitotemporal cortex, a region known to show a preference for small inanimate over large inanimate objects (Konkle and Oliva, 2012; Konkle and Caramazza, 2013). For example, in this territory, there is a region with

preference for tools (Chao et al., 1999; Bracci et al., 2012; Gallivan et al., 2013; Chen et al., 2016) which is closely overlapping with effector-specific regions (Bracci et al., 2012), and persists in the congenitally blind (Peelen et al., 2013; Wang et al., 2015). Further, this cortex shows connectivity with frontoparietal networks supporting actions (Simmons and Martin, 2012; Bi et al., 2015; Konkle and Caramazza, 2017), and is causally involved in tool-action discrimination judgments (Perini et al., 2014). While past studies have typically focused on objects that require hand-performed actions, the present study allows us to test whether the inclusion of non-hand-performed actions through the motor-relevance dimension would further explain neural responses and extend them to large objects.

In the ventromedial temporal cortex, another tool-preferring region has been observed by some studies in the medial fusiform gyrus when contrasting tools to animals (Chao et al., 1999; Whatmough et al., 2002; Mahon et al., 2007; Garcea and Mahon 2014). However, the property of motor-relevance in this region is less clear: for example, responses to manipulable tools are not higher than to other inanimate objects (Mahon et al., 2007; Chen et al., 2018), though there is still some evidence for sensitivity to tools in neural adaptation signals (Mahon et al., 2007). Further, research focused on scene understanding has characterized these ventral-medial object responses along other kinds of object properties that are less related to motor-relevance. For example, responses in ventromedial scene-related regions are best predicted by objects invoking a local space that tend to stay fixed in the world, and are useful as a landmark (Troiani et al., 2014, Mullally and Maguire, 2011; Julian et al., 2016; Auger et al., 2013; see also Epstein, 2014). These same areas are also known to show a preference for large over small inanimate objects (Konkle and Oliva, 2012; Konkle and Caramazza, 2013) and, interestingly, also to the names of such objects in both blind and sighted individuals (He et al., 2013).

Given these past observations, in the present study we aimed to clarify how brain responses to different objects in occipitotemporal cortex reflect motor-relevance and object size properties. To do so, we designed our experiment to enable two different levels of granularity. First, we compared the broad relationship among object size and motor-relevance properties along a 2x2 design (large vs small, motor-relevant vs non-motor relevant), finding effects for both variables independently. Second, we examined brain responses at the level of object categories (72

127 categories total), where a factor analysis revealed that three dimensions that respectively relate to
128 object size, manipulability, and stability interact to account for the structure of brain responses in
129 different regions. Finally, we consider the relative merits of visual feature- versus domain-based
130 properties (e.g., manipulability, navigation relevance) as driving principles in the organization of
131 high-level visual cortex.

2. METHODS

2.1 Participants

For the fMRI experiment, 21 healthy adults with normal or corrected-to-normal vision participated in a 2-hour fMRI session (age, 18-30 years; 11 females). Three were later removed for excessive head movement, and the remaining 18 subjects were analyzed. Each subject gave informed consent according to procedures approved by the Institutional Review Board at the University of Trento.

For the online behavioral experiments, ratings were collected from 644 total participants (464 subjects for the stimulus set development and validation over the dimensions of motor-relevance [n=161], size [n=168] and recognition [n=135]; and 36 subjects for each of the dimensions of manipulability, spatial definition, position variability, motion-based identity and interaction envelope).

2.2 Stimuli.

The stimulus set consisted of images from 72 object categories, where each object category belonged to one of four conditions (small motor-relevant, small non-motor-relevant, large motor-relevant and large non-motor-relevant) with 18 object categories per condition. Each object category included images of six exemplars of that category, for a total of 432 pictures in the stimulus set (Figure 1a). Each image depicted an isolated object on a white background, and the image was sized to 512 x 512 pixels with the object's longest axis reaching the border of the image.

2.2.1 Stimuli validation. To develop and validate the stimulus set, we performed three separate behavioral studies using Amazon Mechanical Turk on an initial larger set of 20 object categories per condition with 7 unique exemplars for each object category (560 total images). Our aim was to construct a stimulus set in which items were recognizable, and where the size and motor-relevance dimensions were balanced.

a. *Recognition ratings.* Observers were presented with an item and reported whether they were familiar with this object (yes/no), and how easily they recognized the object on a scale from 1 (very hard to recognize) to 5 (very easy to recognize).

b. *Motor-relevance ratings.* Participants were presented with an item and rated the degree to

which the object made them think of moving their hands or other parts of their body, using a scale from 1 (not at all) to 7 (very strongly).

c. *Size Ratings.* Participants were presented with an item and judged the size of an object as they would encounter it in everyday life, on a scale from 1 to 7. To provide a reference for the scale, pictures of a wristwatch (small size), a hiking backpack (medium size) and a bed (large size) were provided beside the values of 1, 4 and 7 on the scale.

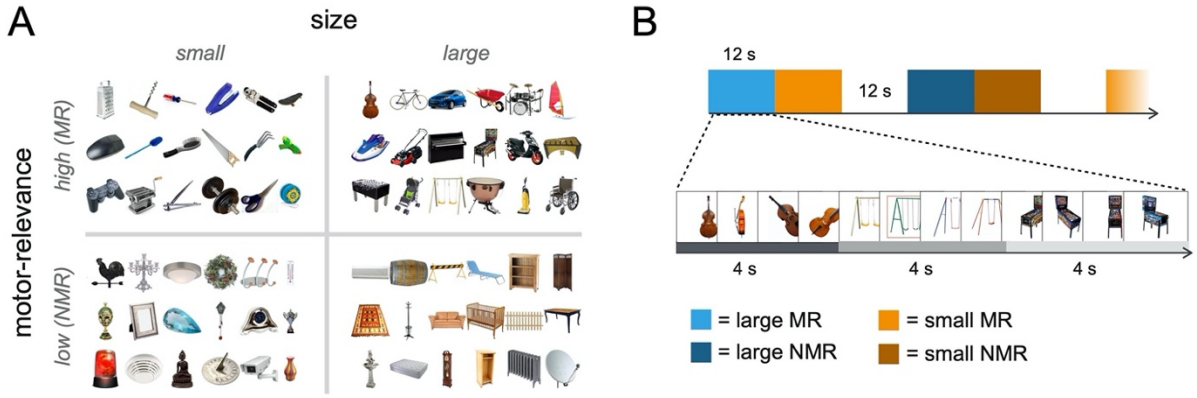
For all three experiments, each participant made judgments for 20 items (5 items from each of the four conditions). Only one exemplar for an object category was presented to each participant. Ratings were obtained for all 560 objects combining data from all participants, so that each image was rated by 6 observers along the measures of size and motor-relevance and by five observers along the measure of recognition.

In order to eliminate those items that were least recognizable, we dropped one item from each object category based on the familiarity score in the recognition experiment. We next eliminated two object categories that scored low in familiarity in order to balance the ratings from the other two experiments (Motor-relevance and Size experiments). This resulted in a new stimulus set with 18 objects per condition (72 total) and 6 items per object. All results and analysis that follow refer to this new stimulus set.

To ensure these factors were balanced, we conducted several statistical tests. As expected, motor-relevant objects had higher motor-relevance scores than non-motor-relevant objects (t_{35}=32.93, $p<0.001$), and large objects had higher size ratings than small objects (t_{35}=19.82, $p<0.001$). Further, large and small objects did not differ in their average motor-relevance scores (t_{35}=1.63, p=0.11). The same was true when looking separately at large motor-relevant and small motor-relevant objects (t_{17}=1.05, p=0.31), and small non-motor-relevant and large non-motor-relevant object categories (t_{35}=1.35, p=0.2). Similarly, motor-relevant and non-motor-relevant objects did not differ in their average size score (t_{35}=-0.72, p=0.5) and neither did large motor-relevant and large non-motor-relevant relevant objects (t_{17}=0.96, p=0.35), nor small motor-relevant and small non-motor-relevant objects (t_{17}=-1.76, p=0.1).

194 **2.3 Image Acquisition.**

195 Imaging data were acquired using a BioSpin MedSpec 4T scanner (Bruker) with an eight-channel
196 coil. Functional data were obtained with an echo-planar 2D imaging sequence (repetition time TR:
197 2000ms; echo time TE: 33ms; flip angle: 73º; slice thickness: 3 mm; gap: 0.99 mm, with 3x3 in-
198 plane resolution and 34 slices). Volumes were acquired in ascending interleaved order of slice
199 acquisition.
200



201
202 *[2 column fitting]*
203 ***Figure 1****. a) Stimulus set: 72 inanimate object categories divided into four conditions (small*
204 *motor-relevant, small non-motor-relevant, large motor-relevant, large non-motor-relevant). Each*
205 *object category included six individual items, for a total of 432 individual images. b) fMRI task:*
206 *subjects were presented with four consecutive pictures of the same object category in a 4s object*
207 *category mini-block. Three 4s category-level mini-blocks (e.g. cello mini-block, swing set mini-*
208 *block, pinball machine mini-block) were presented sequentially in a 12s condition-level main block*
209 *(e.g. large motor-relevant objects block). The task for participants was to detect when a red frame*
210 *surrounded one of the objects, which happened once per 12s block.*
211

2.4 Experimental Design and Statistical Analysis

2.4.1 Design and procedure. During the fMRI experiment subjects viewed images of the small motor-relevant, small non-motor-relevant, large motor-relevant, large non-motor-relevant objects in a classic blocked design. In each run, stimuli from each of the 4 conditions were presented 6 times in 12s stimulus blocks, with 6 rest blocks of 12s interspersed. Images were each presented for 800ms with a 200ms blank between each stimulus. Fixation blocks also appeared at the beginning and end of each run, for 4 and 12 seconds respectively. Participants were instructed to maintain fixation and to press a button when a red frame surrounded one of the objects, which happened once per block. The full experiment consisted of six runs each lasting 6 minutes and 12 seconds.

Additionally, we added sub-structure in each stimulus block, to support more exploratory analyses about the relationship between all 72 object categories (see Figure 1b). Specifically, each stimulus block was comprised of 4s mini-blocks. In each mini-block, three different exemplars from the same category were shown. For example, in a 12s block of large motor-relevant objects, a participant would see 4 different cellos, followed by 4 different swings, followed by 4 different pinball machines. Within each run all 72 object categories were presented exactly one time. The specific 4 exemplars that were displayed in each mini-block were randomly chosen without repetition from the set of 6 possible exemplars from that object category.

Each image was presented in isolation at a ~8 x ~8º visual angle. Stimulus presentation was performed using the Psychophysics Toolbox package (Brainard, 1997) in MATLAB.

2.4.2 Functional Localizer. Three independent functional localizer runs were performed to identify category-selective regions. These runs consisted of a blocked design, in which each block included one of the following stimuli: 1) objects 2) scrambled objects 3) scenes 4) bodies 5) hands. Each run lasted 6 minutes and 36 seconds, with 50 6-second stimulus blocks (10 per condition) and ten 8-second rest blocks interspersed. In each stimulus-block, six different stimuli from the same condition were presented, where each image was shown for 800ms followed by 200ms of fixation. Additionally, there was an orthogonal motion manipulation: for each of the 5 stimulus categories, in half of the blocks the object was presented at the center of the screen, and in the other half the object moved from the center out in one of 8 randomly directions randomly.

Observers were instructed to maintain fixation throughout the experiment and press the button when an exact image repeated back-to-back.

Two of the 21 subjects performed 8 experimental runs (instead of 6) and only 2 Localizer runs (instead of 3). Data from all experimental runs for these subjects were used for data analysis. Due to a technical error, localizer runs for one subject did not include scrambled objects.

For the purpose of the current project, we were only interested in identifying category-selective regions Lateral Occipital Complex (LOC; Objects > Scramble; Grill-Spector, 2003), parahippocampal place area (PPA; Scenes > Objects; Epstein and Kanwisher, 1998) and Occipital Place Area (OPA; Scenes > Objects; Dilks et al., 2013). We were not able to localize area LOC in the subject for whom the scrambled objects condition was missing. To compare the coordinates of our contrasts of interest with the location of well-known category-selective regions, we also localized extrastriate body area (EBA; Downing et al., 2001) and a nearby hand-selective area (Bracci et al., 2012).

2.4.3 fMRI Data Analysis.

Functional neuroimaging data were analyzed using SPM12 (Ashburner et al., 2014), MARSBAR (Brett et al., 2002) and bspmview (https://www.bobspunt.com/software/bspmview/) on MATLAB (Versions 2014b and 2016a,b). Barplots and scatterplots were displayed using ggplot2 (Wickham, 2009) in RStudio.

The raw functional images were submitted to preprocessing, where the first 4 volumes were discarded from each run, slice scan-time correction was performed, followed by 3D motion correction, normalization and spatial smoothing (8 mm FWHM kernel). Data were modeled using standard general linear models (GLM). The first GLM included regressors for each of the four main conditions, with run regressors and motion correction parameters included as nuisance factors. The second GLM modeled each object category as a separate condition for a total of 72 regressors, with run regressors and motion correction parameters included as nuisance factors.

2.4.4 Regions-of-interest selection. We created our ROIs utilizing a separate set of runs from the ones we used for data-analysis. Spherical ROIs were defined around the individual peaks of activation for the whole-brain size contrast Small > Large (collapsing over motor-relevance) from two experimental runs. Univariate responses were extracted from the remaining experimental runs

for size and motor-relevance conditions. These ROIs were defined with a 12mm sphere (257 voxels) centered around the peak positive and negative voxels in individual subjects, and then intersected with individual whole-brain gray matter masks. The selected peaks are referred to following the conventions of Konkle and Oliva (2010) and included small-object preferring bilateral inferior temporal gyrus (Small-ITG, left hemisphere [LH] n=15, right hemisphere [RH] n=17), and large-object preferring bilateral PHC (Large-PHC, LH n=16, RH n=15). After intersecting the 12mm ROI sphere with each subjects' gray-matter mask, the ROIs slightly varied in size across subjects. After intersection, the average size was 244 voxels for Small-ITG and 257 voxels for Large-PHC.

We further identify large-object preferring TOS (Large-TOS) dorsally based on the size contrast, and a series of category-selective regions using the Functional Localizer runs. The regions identified with the Localizer were object-selective lateral occipitotemporal cortex (LOC), scene-selective parahippocampal place area (PPA), and scene-selective occipital-place area (OPA). Analysis and results for Large-TOS and the category-selective ROIs are reported in the Supplementary Material.

2.4.5 Voxel mask. We produced a mask of the most reliable object-selective voxels in occipital, temporal and parietal regions using the reliability-based voxel selection method for condition-rich designs (Tarhan and Konkle, 2020). To compute the reliability maps, average beta values for odd and even runs were separated in two groups and correlated for each voxel that was included within the gray matter. To establish a reliability threshold, we then correlated the patterns for the same object category at different thresholding levels and picked the threshold at which the average correlation across conditions plateaued, which in this case was $r = 0.25$. We then produced a mask from the reliability map at this value.

2.4.6 Whole-brain conjunction analysis. To explore the size organization across motor-relevance, we performed a random-effects conjunction analysis at the individual level between small motor-relevant > large motor-relevant and small non-motor-relevant > large non-motor-relevant.

For each individual, at each voxel, t-values for the two contrasts were compared. The β-weight corresponding to the contrast with t-value closest to zero was then assigned to the given voxel.

This resulted in eighteen conjunction maps (one for each participant) which were then submitted

to a one-sample t-test, and the resulting t-values were assigned to the corresponding voxels on a new map. The voxels with highest value in these new t-maps are those that are strongly activated or deactivated by both contrasts in most subjects. The resulting maps were thresholded with q(FDR)<0.05. Similarly, to explore the motor-relevance organization across size, we performed the same analysis but considering small motor-relevant > small non-motor-relevant and large motor-relevant > large non-motor-relevant contrasts.

2.4.7 Typical-Atypical contrast analysis.

To further explore the relationship between size and motor-relevance, we compared maps between two contrasts: a *typical contrast* [small motor-relevant > large non-motor-relevant], which follows the real-world covariation of small objects being more motor-relevant than large objects; and an *atypical contrast* [small non-motor-relevant > large motor-relevant]. Individual maps for each contrast were submitted to a second-level analysis to obtain a group map, which was FDR-thresholded at q<0.05.

To assess the reliability of the two contrasts, we further produced individual maps for each contrast separately for odd and even runs and correlated them. We then submitted individual correlations for the two separate contrasts to a t-test to assess whether there was a significant different in reliability between the maps produced by the two contrasts.

2.5 Exploratory Analysis.

2.5.1 Object Property Ratings. Ratings for the 72 objects in our dataset were collected on a series of other dimensions to identify the ones that were best associated with the response profiles of our ROIs. All studies were conducted on Amazon Mechanical Turk with 36 judgments per object category. Each of the participants (n=36) rated one exemplar of each of the 72 object categories, with six participants rating the same exemplars within each category. Participants were not presented with two exemplars from the same category. The following object properties were rated:

1. *Manipulability.* Manipulability tracks the degree to which hand-interactions with objects are prominent. Participants were presented with an item and rated the degree to which the object made them think of moving their hands, using a scale from 1 (not at all) to 7 (very strongly). The question asked to participants was very similar to the one posed to collect

the motor-relevance dimension, but here we focus on actions performed with the hands rather than with other parts of the body.

2. *Motion-based Identity*. The degree to which an object's parts are expected to move and change position might be a relevant property for representing objects. Participants were asked to rate on a scale from 1 to 7 the degree to which motion is important in determining the identity of the object.
3. *Spatial Definition*. This property is defined as "the degree to which objects evoke a sense of surrounding space" (Mullally and Maguire, 2011). Participants were asked to rate on a scale from 1 to 7 the degree to which a background is evoked when looking at a specific object.
4. *Position Variability*. Objects can be more or less likely to change location in the environment (i.e., a car is more likely to change location than a swing). Participants were asked to rate on a scale from 1 to 7 the degree to which an object is likely to change position in everyday life.
5. *Interaction Envelop*e. "Interaction envelope" was operationalized as a measure of the amount of space needed to interact with the object (e.g., a hamburger needs to be manipulated with two hands that cover its whole surface, thus it has a larger interaction envelope than a coffee mug which just needs one hand on its handle to be manipulated), following Bainbridge and Oliva (2015). Participants had to judge on a scale from 0 to 2 how many hands are often used when interacting with an object.

2.5.2 Pairwise correlations. We compared the similarity across our seven dimensions (size, motor-relevance and the additional ones collected post-hoc), and the reliability of each of the rated dimensions, in the following way. First, we divided participants into two groups and computed the mean rating for each group and each object category separately for each dimension, resulting in two group vectors of 72 mean ratings for each dimension. To compute the reliability of a dimension, we correlated the two group vectors with each other. To compute the correlation between two different dimensions, we correlated each dimension's group vectors with the other dimension's group vectors, and averaged across the four combinations.

2.5.3 Factor Analysis. To understand how our critical dimensions of size and motor-relevance

related to the variety of stimulus properties from the subsequent stimulus norming studies, we conducted a maximum-likelihood factor analysis on all seven dimensions (size, motor-relevance, manipulability, motion-based identity, spatial definition, position variability and interaction envelope) with a "varimax" rotation of the coefficients. This factor analysis was implemented with the *factanal* function in R. To determine the number of meaningful factors to extract from our dimensions, we performed a "parallel" analysis (Horn, 1965). This procedure involves generating random data and submitting them to the same factor analysis, iterated 5000 times. The average of the eigenvalues resulting from the parallel random-data factor analyses is then compared to the factors from the observed data: if the average eigenvalues from the parallel factor is smaller than the one from the data factor, then the data factor is kept. This procedure was implemented with the *fa.parallel* function in R. The three resulting factors can be summarized as relating to real-world size and context (f_{size}), manipulability (f_{manip}) and degree of movement ($f_{stability}$). For the following ROI and searchlight analyses, for ease of interpretation, we flipped the direction of $f_{stability}$ so as to positively correlate with degree of stability.

2.4.6 ROI analysis with Factors as predictors. Following the factor analysis on the behavioral ratings of the objects, we performed a linear model analysis to predict the size ROIs responses from the resulting factor scores and their interactions. To estimate the response profile of each ROI, β-weights for each of the 72 object categories were extracted from the GLMs of the main experiment runs. Note that this analysis was performed on runs separate from the ones used to produce the ROIs. These betas were averaged across voxels, resulting in a vector of 72 values reflecting the ROIs overall response activation to each category. The ROIs activation profiles were computed for each subject using subject-specific ROIs and GLMs.

We next sought to explain these ROI activation profiles with the factors from the factor analysis above by employing a linear modeling approach. All linear models to predict an ROI activation were implemented with the function *lm* in R. We first looked for hemispheric interactions by submitting our size ROIs to a model with the three factors, hemisphere and their interaction as coefficients. No hemispheric interaction significantly explained either Small-ITG or Large-PHC's activation; hemisphere was thus removed from the models.

Next, we examined which of three nested models best explained activation in our size ROIs by comparing their resulting adjusted R square (adjR^2). The three models were: one with the three factors as predictors, with no interaction terms [$f_{size} + f_{manip} + f_{stability}$]; a second model that also included the two-way interaction terms [$f_{size}*f_{manip} + f_{size}*f_{stability} + f_{manip} *f_{stability}$]; and a third model that also included the three-way interaction term [$f_{size}*f_{manip}*f_{stability}$]. The model with the highest adjR^2 was considered the one most fitting the region and the significant coefficients were explored. If two models had comparable adjR^2 (rounded at the second decimal), the simpler model was picked. The same analysis was also performed on Large-TOS and in category-selective regions defined from the Functional Localizer (LOC-sphere, PPA-sphere and OPA-sphere; see Supplementary Material).

When comparing the nested models for the highest adj R^2, we found that both Small-ITG and Large-PHC were best explained by a model with all three factors and their two-way interactions (Small-ITG: adj. $R^2 = 0.33$; $F_{6,65} = 6.90$, $p < 0.001$; Large-PHC: adj. $R^2 = 0.27$; $F_{6,65} = 5.36$, $p < 0.001$), as compared to the model with only main effects (Small-ITG: adj. $R^2 = 0.23$; Large-PHC: adj. $R^2 = 0.21$) and the model with main effects, two- and three-way interactions (Small-ITG: adj. $R^2 = 0.32$; Large-PHC: adj. $R^2 = 0.26$).

To evaluate the quality of the fit of the different coefficients, we looked at the resulting associated t- and p-values. T-values were computed by dividing coefficients by their standard errors, while p-values tested the significance against the hypothesis of obtaining given t-values if the coefficients were not actually contributing in explaining the dependent variable.

2.5.4 Searchlight Regression Analysis. Following our findings in the spherical ROIs, we next explored which factors would best fit voxels throughout object-responsive cortex.

The searchlight analysis was performed in the following way: within the voxel mask, a 6mm sphere was drawn around a voxel, and all voxels falling within that sphere were considered part of the neighborhood for that voxel. This procedure was performed iteratively so that each voxel within the object-selective mask was at the center of the sphere. In each searchlight ROI, voxel activation for all 72 object categories was averaged across all voxels within the sphere, and the resulting

vector used as the dependent variable for two models.

For each search sphere, we fit two models. The first model related size and manipulability, with f_{size}, f_{manip}, and their interaction [$f_{size} + f_{manip} + f_{size} * f_{manip}$], which we refer to as *manip-model*. the second model related size and stability, with f_{size}, $f_{stability}$ and their interaction [$f_{size} + f_{stability} + f_{size} * f_{stability}$], which we refer to as *stability-model*. For each given searchlight, adjusted R^2 values for these two models were extracted and compared, and the model with the highest positive value was selected and assigned to that searchlight's central voxel.

To visualize which models are fitting best and where, we looked separately at small-preferring and large-object preferring voxels. Small- and large-object preferring voxels were identified based on whether the coefficient for the main effect of f_{size} was positive (preference for small) or negative (preference for large). We then visualized the coefficient maps for the interaction terms $f_{size}*f_{manip}$ and $f_{size}*fs_{tability}$ separately for large- and small-preferring voxels, comparing them with a map of the F-values for the winning model associated with each voxel.

3. RESULTS

3.1 Whole-brain conjunction analysis

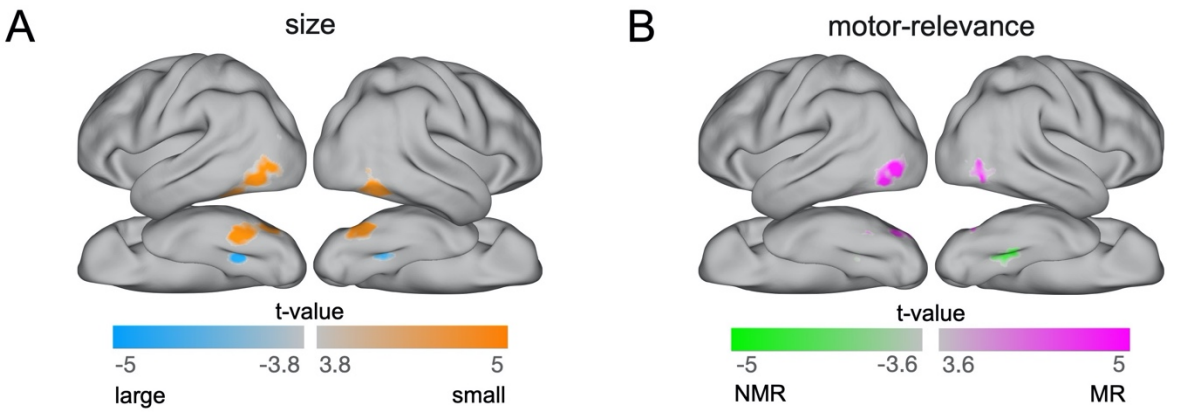
We first examined whether the large-scale neural organization by real-world size of inanimate objects is preserved when disrupting the natural covariance between size and motor-relevance. To do so, we conducted a conjunction analysis to isolate regions which show a large vs small object difference that holds when comparing motor-relevant as well as non-motor-relevant objects (see Methods).

When looking at the real-world size organization, we found regions that showed a preference for small over large objects, and for large over small objects, regardless of the motor-relevance preference. The whole brain maps can be visualized in Figure 2a, with the group MNI coordinates at the bottom of the figure. With respect to small objects, we observed activation bilaterally in ITG and in the left hemisphere and along the lateral aspect of the fusiform gyrus. In the left hemisphere, the region for small objects expanded posteriorly to the lateral occipital cortex. With respect to large objects, along the ventral surface of the occipitotemporal cortex, we observed bilateral regions corresponding to the parahippocampal cortex. A preference for large objects was also present more dorsally in the lateral surface in the occipital place area (OPA) in individual participants, but this region did not survive the group-level conjunction test at the FDR threshold depicted in Figure 2a. Overall, this group-level conjunction analysis reveals that, even when taking into account motor-relevance, the real-world size of objects drives large-scale differential responses across the ventral stream.

Next, we examined whether also motor-relevance, when disrupting its natural covariance with size, would elicit a large-scale organization across this cortex. Another whole-brain conjunction analysis was conducted, comparing motor-relevant objects vs non-motor-relevant objects, requiring this relationship to hold for both large object and small objects independently. The results are mapped in Figure 2b, with the group MNI coordinates at the bottom of the figure. Along the lateral surface, we observed regions with a stronger response to motor-relevant objects, with peaks in LO bilaterally. Along the medial aspect of the ventral surface, we found stronger responses to non-motor-relevant objects primarily in the right hemisphere, with the peak activation located slightly anterior to the large-object preferring region. This analysis reveals that the

dimension of motor-relevance also elicits stable response differences in parts of object-responsive cortex.

Notably, these conjunction maps yielded relatively similar organizations. Specifically, lateral regions prefer both small and motor-relevant objects; while ventromedial regions prefer both large and non-motor-relevant objects. These findings indicate that neither size alone nor motor-relevance alone are sufficient to explain the large-scale organization of responses to objects along occipitotemporal cortex. Further, these results highlight that there is a clear association between the two dimensions and how they map across the cortex.



Conjunction Contrasts	x	y	z	Peak t-value
small: *small MR > large MR & small NMR > large NMR*				
L ITG	-40	-66	-8	8.28
L Fus	-42	-44	-18	7.46
L LO	-40	-88	-6	4.26
R ITG	46	-56	-10	7.41
large: *large MR > small MR & large NMR > small NMR*				
L PHC	-28	-50	-2	4.64
R PHC	32	-46	-6	-5.28

Conjunction Contrasts	x	y	z	Peak t-value
MR: *small MR > small NMR & large MR > large NMR*				
L LO	-46	-82	0	7.64
R LO	42	-70	2	5.65
NMR: *small NMR > small MR & large NMR > large MR*				
L PHC	-26	-44	-14	-5.80
R PHC	28	-36	-18	-5.71

[2 columns fitting]

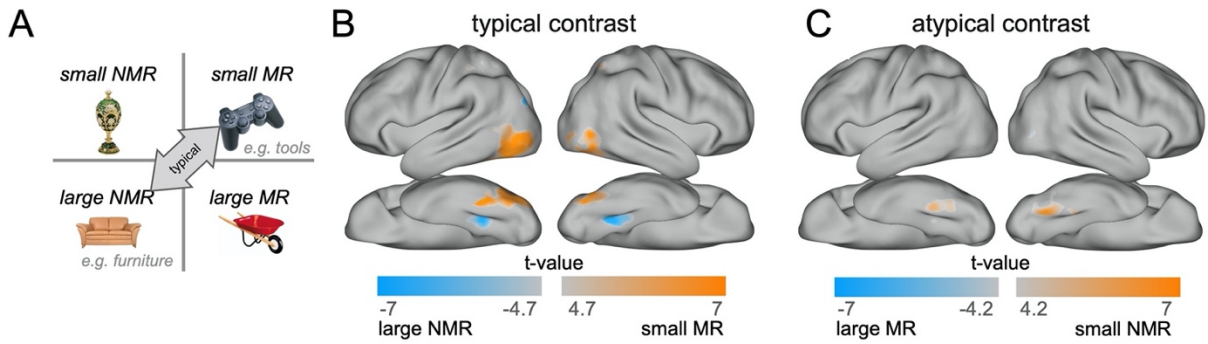
Figure 2. Whole Brain Random-Effects Conjunction Analysis. *a) Size conjunction contrasts, where both [small motor-relevant > large motor-relevant] and [small non-motor-relevant > large non-motor-relevant] relationships hold. a) Motor-relevance conjunction contrasts, where both [small motor-relevant > small non-motor-relevant] and [large motor-relevant > large non-motor relevant] relationships hold. Contrasts were thresholded at q(FDR)<0.05.*

3.2 Typical-atypical Contrasts

To further explore the relationship between size and motor-relevance, we compared maps between two contrasts: a *typical contrast* (small motor-relevant > large non-motor-relevant), which follows the real-world covariation of small objects being more motor-relevant than large objects; and an *atypical contrast* (small non-motor-relevant > large motor-relevant; Figure 3a). These maps are shown in Figure 3.

When looking at surface maps produced by the typical contrast, we find extensive activations along the lateral and ventral surfaces (Figure 3b). In contrast, the surface maps produced by the atypical contrast (Figure 3c) are less extensive and also less reliable in a split-half analysis (typical split-half map correlations: M=0.61, SD=0.18; atypical: M=0.43, SD=0.24; t_{17}=2.70, $p<0.05$).

These topographic observations mirror the ecological covariation of these factors (Figure 3a): while it is theoretically possible to artificially dissociate size and motor-relevance as we have done in this stimulus set, in every day experience small objects tend to be more motor-relevant and large objects less so. And we found this covariation to also be evident in the neural data (see Figure 3b,c). Taken together, these data show the importance of the interaction between size and motor-relevance in driving more reliable and robust neural responses, thus revealing the importance of the natural covariation between these two properties in the visual system.



[2 columns fitting]

Figure 3. Typical/Atypical contrast Analysis. *a) simplified representation of the size x motor-relevance design. Arrow points to the typical contrast conditions, representing the covariation commonly observed in the environment. b) Typical contrast: small motor-relevant > large non-motor-relevant. c) Atypical contrast: small non-motor-relevant > large motor-relevant. Contrasts were thresholded at q(FDR)<0.05.*

3.3 Exploratory Analyses

3.3.1 Factor Analysis. We next explored the possibility that the relationship between size and motor-relevance may be more multifaceted than what can be gathered from a simple 2 x 2 design. Our goal for this analysis was to understand and explore how other, related properties that have been proposed in the past relate to our findings and predict neural responses at the category level. To do so, we collected ratings along five additional dimensions of our 72 object categories and submitted them to a factor analysis together with size and motor-relevance.

First, manipulability ratings were obtained, which focused on hand movements specifically, in order to understand the extent to which our motor-relevance dimension correlated with this previously studied dimension (Mahon et al., 2007). We also examined a number of dimensions that have been previously proposed as significant modulators of tool regions or scene regions, namely interaction envelope (Bainbridge and Oliva, 2015), spatial definition (Mullally and Maguire, 2011), and motion-based identity (Beauchamp, 2005). Finally, based on our observation regarding low responses to vehicles in Large-PHC, we obtained "position variability" ratings, capturing how often an object changes its position in the environment. Note that these properties were defined after having seen the data, and thus any role they play in explaining neural data requires independent follow-up studies to confirm.

The relationship among these ratings and our original size and motor-relevant dimensions are detailed in Figure 4. All ratings had a relatively high reliability (Figure 4a, shaded diagonal numbers), with a range of similarity relationship between any pair of dimensions (Figure 6b). There are a couple observations to note from inspecting these pairwise correlations and scatter plots. First, by design, size and motor-relevance were completely uncorrelated in this stimulus set. Second, the correlation between motor-relevance and manipulability was at ceiling, virtually identifying the same property. As mentioned in the introduction, most motor-relevant objects involved hands for their interaction; including whole-body actions in the motor-relevance score had a minimal or non-existent effect (the only exception being the skateboard which was high in motor-relevance but low on manipulability).

Next, we extracted the latent dimensions within this dataset using factor analysis (see Methods). A parallel analysis procedure determined that three factors should be retained, which were sufficient to explain more than 80% of all variance among the ratings. The property loadings for each factor are shown to the right in Figure 4a.

The first factor was loaded mostly by manipulability, motor-relevance and interaction envelope—it seemed thus to be most related to *actions and hand-object interactions*. The second factor was loaded by size and spatial definition, and was thus most related to the *physical presence* of the object. Finally, motion-based identity, position variability and motor-relevance loaded strongly on the third factor, which seemed thus to capture the *mobility/movability of the object*.

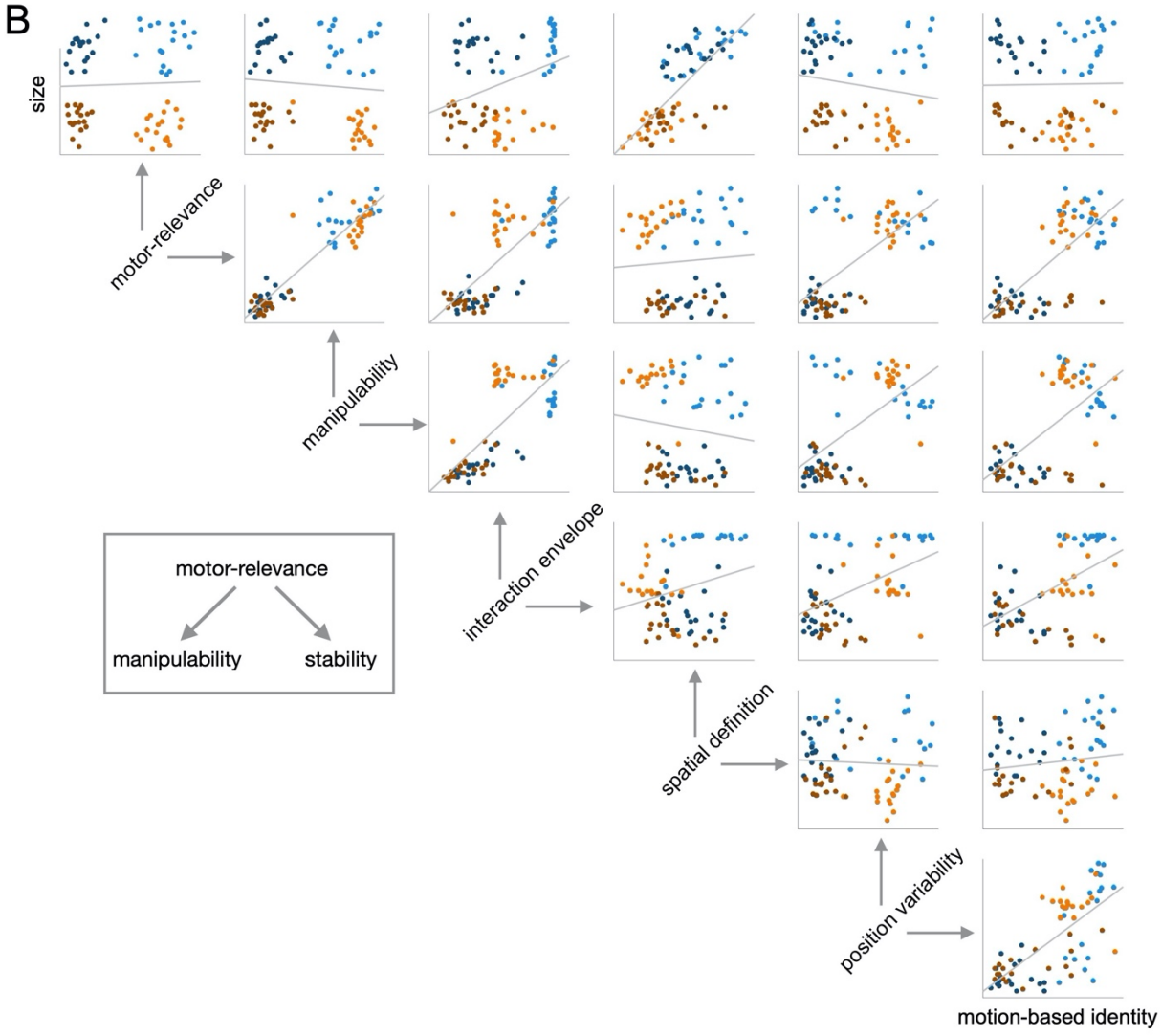
Considering these factors within the context of our main design, our motor-relevance dimension was effectively split into two factors, one related to hand-object interactions (f_{manip}) and one related to an object's mobility/movability ($f_{stability}$), and both of these were separate from the more physical properties of size and, to a lesser degree, spatial definition (f_{size}). Thus, while our stimulus set was collected as varying along only two primary factors (size and motor-relevance), the factor analysis indicated that this stimulus set is better characterized along three factors (size, manipulability and stability).

We next explored how well the three factors extracted from our ratings can explain the neural response profiles, both in a targeted ROI analysis and in a broader searchlight analysis. Note that these analyses were unplanned and should be considered post-hoc and exploratory in nature.

A

Dimension	Question	Pairwise Correlations							Factor Analysis		
		size	MR	man	IntEnv	SD	PV	MBI	Var	f_{manip} 34%	f_{size} 25%	f_{stab} 24%
size	How big is this object in every day life?	0.99	0.03	-0.09	0.35	0.72	-0.16	0.03		0.12	**0.95**	-0.15
motor-relevance	Interactions require moving hands/body?	0.03	0.97	**0.92**	0.76	0.06	0.66	0.73		**0.81**	0.02	**0.53**
manipulability	Particular hand movements involved?	-0.09	**0.92**	0.97	0.67	-0.18	0.67	0.66		**0.91**	-0.15	0.37
interaction envelope	How many hands are used?	0.35	0.76	0.67	0.97	0.24	0.47	0.55		**0.77**	0.31	0.29
spatial definition	Is a background evoked?	0.72	0.06	-0.18	0.24	0.82	-0.06	0.15		-0.04	**0.83**	0.11
position variability	Change position in the environment?	-0.16	0.67	0.67	0.47	-0.06	0.97	0.69		0.38	-0.11	**0.70**
motion-based identity	Motion important in identifying object?	0.03	0.73	0.66	0.55	0.15	0.69	0.92		0.39	0.09	**0.82**

B



[2 column fitting]

Figure 4. Object dimension similarity and Factor Analysis. *a) For each rated object property, the subplot shows the dimension label, the question asked to elicit ratings on the dimension, the correlation to the other dimensions, where the rating's reliability is shaded, and the dimension's loading in a factor analysis. b) Scatter plots are shown for all pairs of 7 dimensions, where each*

dot reflects an object category and is color-coded based on its size and motor-relevance rating (light orange – small motor-relevant; dark orange: small non-motor relevant; light blue: large motor-relevant; dark blue: large non-motor-relevant).

3.3.2 ROI Analysis. To explore how well the three factors could explain fMRI responses, we employed a linear modeling analysis explaining the activation profile in each ROI, with the three factors obtained from the factor analysis and their interactions as regressors (see Methods). The goal was to see which weighted combination of factors best predicted a region's response variation to the 72 objects. This finer-grained analysis looking at responses to all 72 object categories was possible due to our nested fMRI protocol design (see Methods), in which the neural responses could be modeled at the category-level for each of the 72 categories independently (see Figure 1). Note that these three factors were determined only from the behavioral ratings of our stimulus set. Further, for interpretive clarity, we reversed the sign of $f_{stability}$ so that high values indicate high "fixedness" (rather than high mobility.) This transformation does not affect the significance of the statistical results.

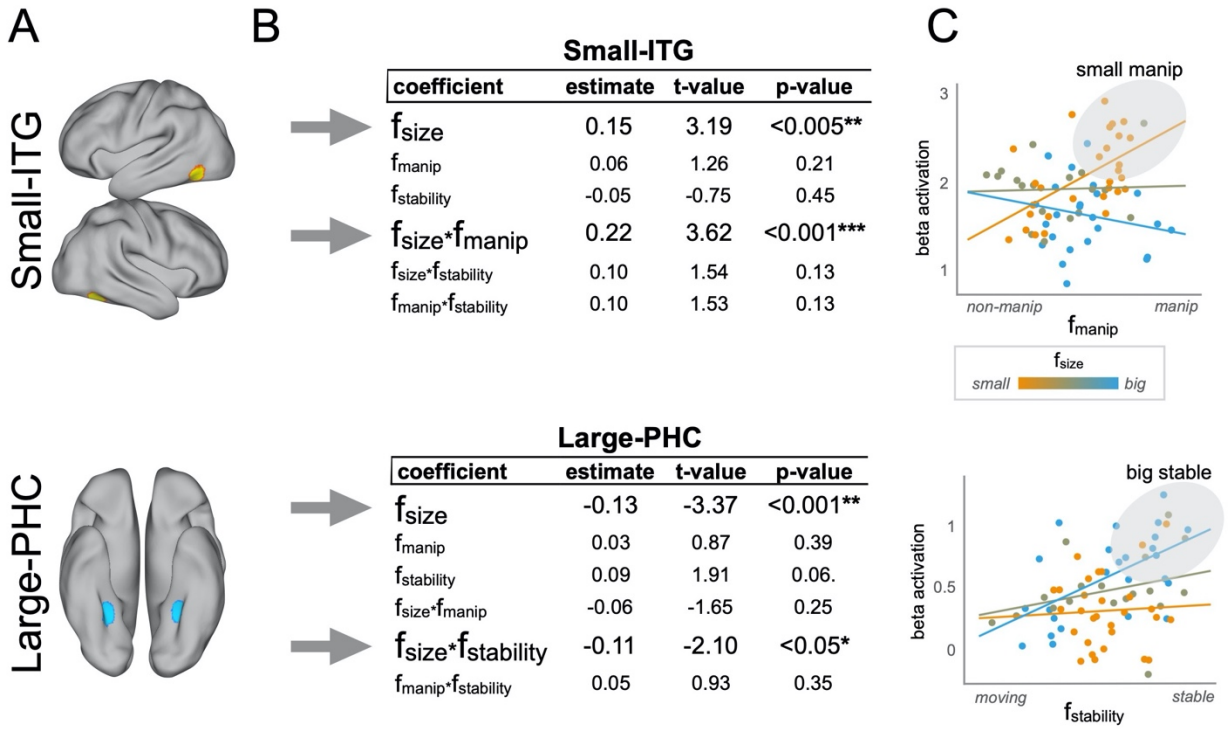
After determining which model was a best fit for each region, we looked at the factors within the winning models that significantly contributed to the fit (see *Methods*), and whether they were positive or negative, to understand the models' relationship to the data (Figure 5a).

In Small-ITG, the results indicate that the best prediction of neural response magnitude was given by a combination of *size* and of the *manipulability* component of motor-relevance: there was a significant effect of f_{size}, as well as the interaction of f_{size} and f_{manip} (see Figure 5a, top for statistical results). The interaction between these factors further indicates that it is specifically small objects involved in hand-object actions which drive the strongest responses in this region. Thus, the distinction of hand-object action and manipulability is a better descriptor of the responses in this region over the more general concept of motor-relevance.

In Large-PHC we observed a different pattern: the coefficients that significantly explained the model were f_{size} and the interaction of f_{size} and $f_{stability}$ (see Figure 5b, bottom row for statistical results). Thus, responses in this region were driven most by large objects that are stable in the environment. This result in Large-PHC clarifies what we observed in earlier analyses employing

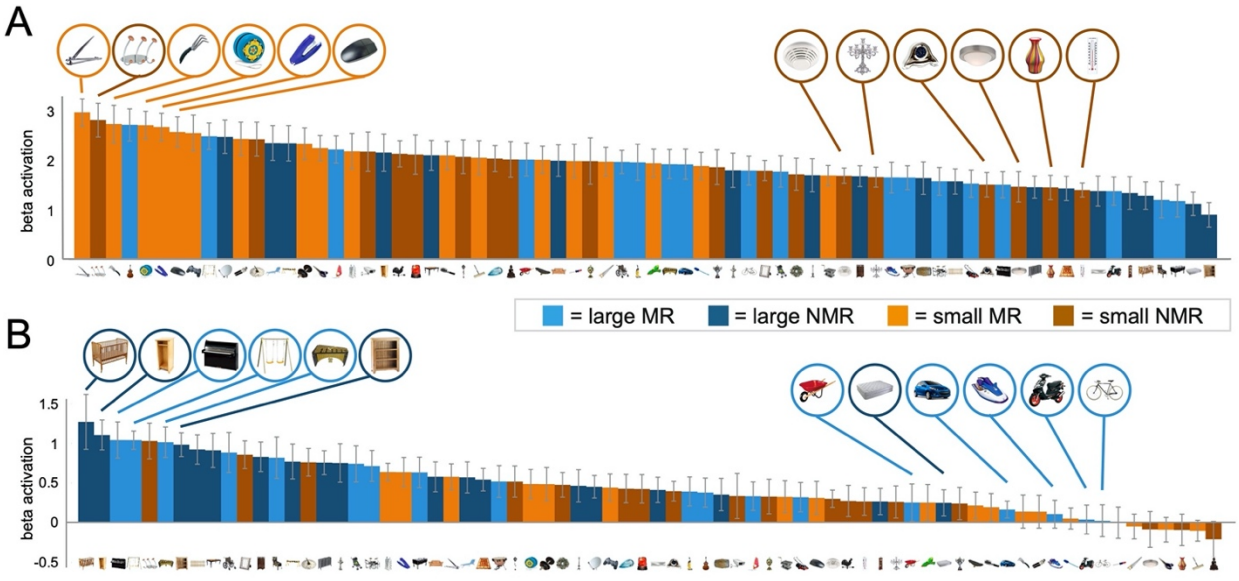
609 the 2 x 2 design: rather than the low-action component of motor-relevance, it might be the *fixedness*
610 component of this construct the one most driving neural representations in this area. That is, the
611 *stability* component of non-motor-relevance is the one to best predict response magnitude in Large-
612 PHC.

613 To better understand these relationships, we visualize them in Figure 5c: for each ROI, the
614 activation to each category is plotted on the y-axis as a function of the manipulability factor for
615 Small-ITG, and the stability factor for Large-PHC. Each dot is color-coded based on the f_{size} value
616 for the corresponding object. This plot reveals that the manipulability factor drives activation in
617 Small-ITG but mostly for small objects (increasing activity for more manipulable small objects,
618 see Figure 5c, top). In contrast, inspection of the scatterplot for Large-PHC suggests that activation
619 is driven by stability but mostly for large objects (increasing activity for more stable large objects,
620 see Figure 5c, bottom). These relationships are also evident in the activation profiles of the two
621 ROIs in Figure 6: For Small-ITG, tools are the small objects that drive the most activation (e.g.
622 nail clipper, hand rack, yo-yo; see Figure 6a), while for Large-PHC, means of transport tend to be
623 the large objects that drive the least activation (e.g. bike, scooter, jet-ski; see Figure 6b).
624



[2 column fitting]

Figure 5 ROI analysis. *a) Visualization of size ROIs Small-ITG and Large-PHC in one example subject. b) GLMs with ROIs' activity as dependent variable and factor scores as independent variables. Shown are estimates for all predictors included in the model, t-value and p-value. The t-value was measured by dividing the coefficients by their standard errors. The p-value tested the hypothesis of obtaining the observed t-value if the coefficient were actually zero.c) Top: scatterplot of Small-ITG's activation by f_{manip} (x-axis) and f_{size} (y-axis). Dots are color-coded based on the f_{size} weight from blue (low-weight, large size) to orange (high weight, small size). Bottom: scatterplot of Large-PHC's activation by $f_{stability}$ (x-axis) and f_{size} (y-axis). Same color-coding scheme as described for the plot above.*



[2 column fitting]

Figure 6 Barplots of ROIs activation. *Mean activation to each of the 72 object categories are plotted for each region (a: Small-ITG; b: Large-PHC), where the categories are rank-ordered from most to least active. Object categories are color-coded based on their size and motor-relevance conditions (light orange: small motor-relevant; dark orange: small non-motor-relevant; light blue: large motor-relevant; dark blue: large non-motor-relevant). Error bars reflect ±1 SEM across subjects.*

Taken together, these analyses begin to characterize the nature of the ecological covariation between size and motor-relevance, clarifying how these object properties fit together. For example, no one factor alone (size, manipulability, stability) best accounted for the responses in any ROI. Rather, manipulability (or hand-relevance) of an object matters more for small objects in the lateral cortex, and stability of an object matters more for large objects in ventral cortex. Finally, these findings are consistent with prior literature that has focused on either the ventral or lateral surface separately (e.g., Bracci et al., 2012; Mullaly and Maguire, 2011), unifying these separate observations in one common data set and analysis procedure.

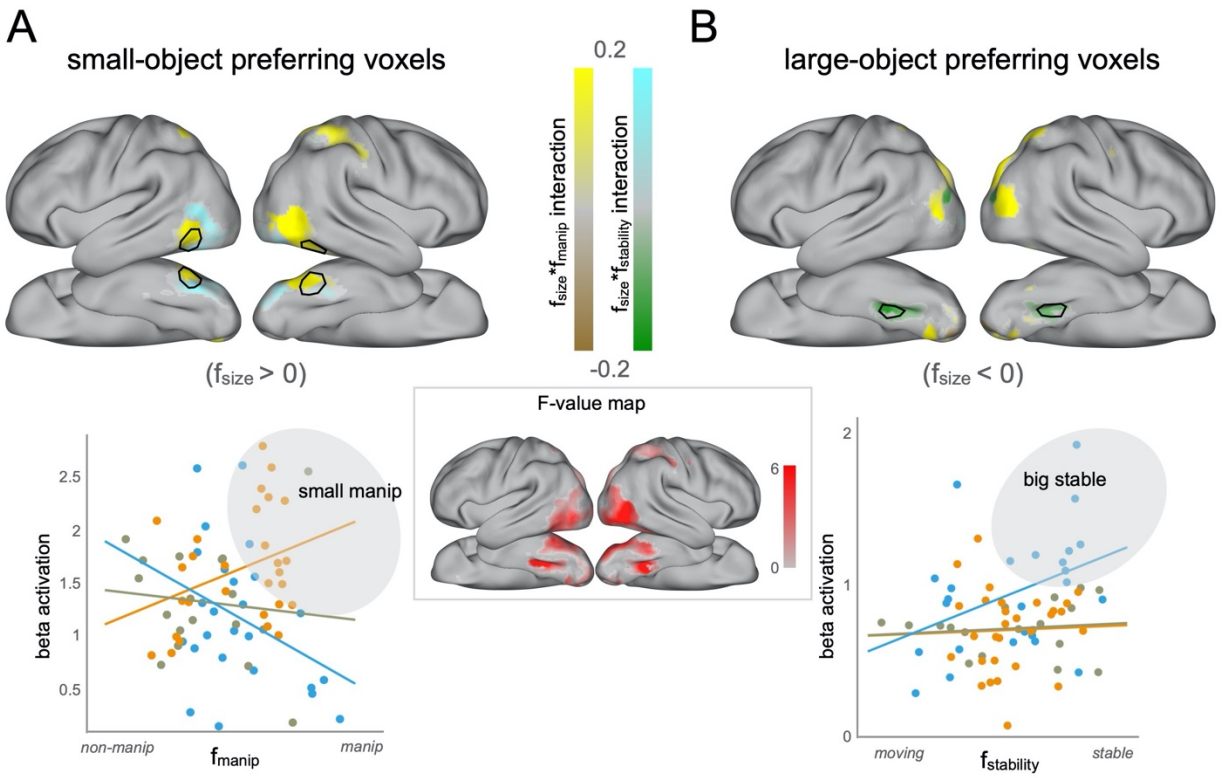
3.3.3 Additional ROIs. To more directly relate our findings to previous literature, we also conducted these same analyses in the classic scene regions, parahippocampal place area and occipital place area (PPA and OPA; Epstein et al., 2014), and object-responsive lateral occipitotemporal cortex (LOC; see Methods and Supplementary Material). These regions are adjacent and partially overlapping with the size ROIs (see also Konkle and Oliva, 2012). The findings in these classic regions generally converge with what was observed for the size ROIs: in LOC, there was a significant contribution of f_{size}, and a significant contribution of the interaction between f_{size} and f_{manip}. In PPA, there was a significant contribution of the size factor, and marginally of stability and of the interaction between size and stability (see Supplementary Material).

Further, we conducted these analyses in an additional occipitodorsal region identified with the [Large > Small] size contrast from two runs of the main experiment (Large-TOS) and in adjacent scene-selective occipital place area (OPA; see Supplementary Material). In Large-TOS, we found a result parallel to our observation in PPA: a significant contribution of f_{size} and of the interaction between f_{size} and $f_{stability}$. However, this result did not replicate in OPA where only f_{size} had a significant contribution.

3.3.4 Searchlight Analysis. By focusing on the ROIs produced by the size contrast, we were able to explore whether other factors beyond size and motor-relevance were important in these ROIs—to which the answer is clearly yes: motor-relevance was indeed split into two sub-dimensions, one more related to manipulability, and one more related to stability, which dissociated in explaining small- and large-object ROIs. However, by performing a targeted ROI analysis, we might have missed additional regions with different associations between these factors.

Given the complexity of the number of possible interactions among these factors, we approached the searchlight analysis with a more targeted comparison. Specifically, we mapped which regions showed a better model fit among two candidate models: one that associates size and manipulability (manip-model) and one that associates size and stability (stability-model; see

687 Methods). All searchlight analyses were conducted within a reliable voxel mask (see Methods).
688 The results of this analysis are shown in Figure 7.
689
690 From this analysis, we can gather two main conclusions. First, we found that the voxels with
691 best fitting models (i.e., the ones with highest F-value, see Figure 7, center) focused mostly in
692 the lateral occipitotemporal region and in the ventromedial occipitotemporal regions, in areas
693 mostly overlapping with size ROIs Small-ITG and Large-PHC (outlined in black in Figure 7a,b).
694 Thus, our targeted ROIs seemed to cover most OTC areas where voxels' information could be
695 explained by a combination of the three factors.
696
697 Second, we found that the best-fitting voxels resulting from the searchlight analysis largely
698 confirmed our targeted ROI analyses. Indeed, small-object preferring voxels in the lateral
699 occipitotemporal region (i.e., the ones with positive f_{size} coefficients for the winning model)
700 presented a positive interaction of the size and manipulability factors (Figure 7a top, colored in
701 yellow), similarly to what was observed in the ROI analysis for Small-ITG: that is, the smaller f_{size}
702 (the smaller the object), and the higher f_{manip} (the more manipulable the object), the stronger the
703 activation. Similarly, we found that large-object preferring voxels in ventromedial
704 occipitotemporal region (i.e., the ones with negative f_{size} coefficients for the winning model)
705 presented a negative interaction of the size and stability factors (Figure 7b, top, colored in green),
706 similarly to what observed in our ROI analysis for Large-PHC: that is, the lower f_{size} (the bigger
707 the object), and the higher $f_{stability}$ (the more stable the object), the stronger the activation. For ease
708 of comparison, we have drawn a black outline on the searchlight coefficient maps of the size ROIs
709 Small-ITG (Figure 7a) and Large-PHC (Figure 7b) for one example subject, from which it is clear
710 that the targeted ROI analyses do in fact highlight the major structure in neural responses by our
711 three factors.
712



713

714 *[2 column fitting]*

715 ***Figure 7. Searchlight Analysis.*** *a) Top: surface map of interaction coefficients for small-object*
716 *preferring voxels (i.e., voxels with a positive f_{size} coefficient). In voxels in which the winning model*
717 *was the manip-model, that voxel is colored according to its f_{size}*f_{manip} interaction coefficient*
718 *estimate in a brown to yellow scale. In voxels in which the winning model was the stability-model,*
719 *that voxel is colored according to its f_{size}*$f_{stability}$ interaction coefficient estimate in a green-to-cyan*
720 *scale. Bottom: scatterplot of activation by f_{manip} (x-axis) and f_{size} (y-axis) for all voxels showing a*
721 *positive f_{size}*f_{manip} interaction in lateral OTC. Dots are color-coded based on the f_{size} weight from*
722 *blue (low-weight, large size) to orange (high weight, small size). b) Top: surface map of interaction*
723 *coefficients for large-object preferring voxels (i.e., voxels with a negative f_{size} coefficient). In voxels*
724 *in which the winning model was the manip-model, that voxel is colored according to its f_{size}*f_{manip}*
725 *interaction coefficient term in a brown-to-yellow scale. In voxels in which the winning model was*
726 *the stability-model, that voxel is colored according to its f_{size}*$f_{stability}$ interaction coefficient estimate*
727 *in a green-to-cyan scale. Bottom: scatterplot of activation by $f_{stability}$ (x-axis) and f_{size} (y-axis) for*
728 *all voxels showing a negative f_{size}*f_{manip} interaction in ventromedial OTC. Dots are color-coded*
729 *based on the f_{size} weight from blue (low-weight, large size) to orange (high weight, small size).*

730 At lower F-value levels, we observed two more nuanced relationships. However, given the
731 lower fit of either manip-model or stability-model within these voxels, one can only draw tentative
732 inferences from these results.
733
734 First, a significant portion of small-object preferring voxels in the lateral occipitotemporal
735 surface (Figure 7a, top, colored in cyan) are fit best by models with size and stability, and show a
736 negative interaction between the size and stability factors. Inspection of activation to all 72 object
737 categories in this area shows a preference for moving over non-moving things (see Supplementary
738 material) which is in line with the presence of movement-preferring region MT+ in the lateral
739 surface.
740
741 Second, we also observed large-object preferring voxels in the dorsal occipitoparietal region,
742 where most voxels (Figure 7b, top, colored in yellow) indicate a positive interaction of the size
743 and manipulability factors while a few voxels (colored in green) indicate a negative interaction of
744 the size and stability factors. However, inspection of the scatter plots for these regions showed that
745 these interactions were not as simple and easily interpretable as the ones observed in areas
746 neighboring our size ROIs (see Supplementary Material). From these results, it is clear that there
747 are systematically varying object responses in this dorsal occipital part of the cortex, but none of
748 the three factors or their interactions cleanly provides a description of what explains the object
749 activations.
750

4. DISCUSSION

Here, we examined whether the large-scale organization of neural responses to object size could be better explained by the factor of motor-relevance. This was not the case – we found that the contrasts of small vs large objects elicited a clear size organization for both motor-relevant and non-motor relevant objects. We also found that motor-relevance, independent of size, accounted for some of the large-scale structure of neural responses. This observation confirms the importance of motor interaction as an explanatory dimension in the ventral visual stream. Crucially, however, further analyses showed the two dimensions to not be independent, in that more typical combinations (small and motor-relevant, large and non-motor-relevant) elicited more reliable representations than less typical ones.

Our subsequent exploratory analyses helped to refine and clarify these results. Specifically, a factor analysis revealed that the construct of motor-relevance could be divided into two underlying dimensions, one better related to *manipulability* (i.e., related to hand performed actions), and one better related to an object's *physical presence* and stability. Further ROI and searchlight exploratory analyses revealed that the manipulability factor was a better descriptor of activation in small-object preferring lateral OTC, while the stability factor was a better descriptor of activation in large-object preferring ventromedial OTC.

Broadly, these results begin to characterize how the ecological covariation between size and behaviorally relevant object properties contributes to the observed dissociations between object categories in the ventral stream regions with a preference for the inanimate domain. In the next two sections, we situate these findings in the literature, and discuss how they inform the deeper theoretical question of what drives the neural organization of inanimate objects.

4.1 Dissociations within the inanimate-domain

Objects are characterized by numerous visual properties, real-world size being one of them. However, size tends to correlate with other visual and non-visual properties. There are suites of partially related factors that predict neural responses to inanimate objects, with a major division between small, manipulable objects on one side and large, stable, navigationally relevant objects on the other (e.g., He et al., 2013; Peelen et al., 2013; Julian et al., 2016; Mullally and Maguire 2011; Troiani et al., 2014; Auger et al., 2013; Bracci et al., 2012; MacDonald and Culham, 2015;

782 Bainbridge and Oliva, 2015). Overall, these findings point to a richer organization than simply
783 object size as such, where different brain regions may be sensitive to size in the presence of other
784 object properties for different reasons (Bi et al., 2015).
785
786 For example, along lateral OTC and middle temporal gyrus, there is a mosaic of partially
787 overlapping regions whose response magnitudes are predicted by a range of related factors, from
788 more primitive object shape (Grill-Spector, 2003), to real-world size (Konkle and Oliva, 2012;
789 Julian et al., 2016), to tools and other objects that are manipulated by the hands (Lewis 2006;
790 Bracci et al., 2012; MacDonald et al., 2015; Mruczek et al., 2013; Lingnau and Downing, 2015).
791 These regions are also adjacent to motion-selective cortex and body-preferring regions (Grill-
792 Spector and Weiner, 2014), including a region that jointly responds to both tools and hands,
793 spanning the animate/inanimate divide (Bracci and Peelen, 2013; Striem-Amit et al., 2017). While
794 the number of regions and exact functional divisions along this cortex are still being clarified, from
795 our results and these others, it is clear that these regions are most strongly related to small,
796 movable, manipulable items.
797
798 Along ventromedial OTC, where scene-preferring PPA is located, there are now a number of
799 converging results characterizing the object properties that are best associated with responses to
800 isolated objects. These factors include the degree to which the object invokes a local tridimensional
801 space, its size, its tendency to stay fixed in the world, and can be generally summarized by an
802 overarching construct of spatial stability and landmark suitability (Troiani et al., 2014, Mullally
803 and Maguire, 2011; Julian et al., 2016; Auger et al., 2013; see also Epstein et al., 2014). While
804 these studies examining PPA responses differ in their stimulus sets, designs, and even mode of
805 presentation (e.g., pictures, mental imagery), they all leverage condition-rich designs and factor
806 analyses to characterize the configuration of related object properties. While our observation of a
807 role of spatial stability in PHC cannot in isolation lead to strong inferences, given that it is the
808 product of a post-hoc analysis, it joins this collection of studies, adding to the diversity of stimulus
809 sets and neuroimaging designs that show converging evidence that size in the context of position
810 fixedness best explains the object responses in this region.
811

Our results, jointly with these collections of studies, suggest that a division based on manipulability for small objects and stability for large objects is fundamental in the organization of inanimate object responses.

4.2 What drives this object organization?

Turning to the broader question of *what drives* the spatial organization of object domains in visual cortex, as opposed to what visual properties are explicitly encoded in domain-specific regions, one answer is that the observed specialization is the result of evolutionary pressures to maximize the efficient mapping of visual representations onto the appropriate downstream regions engaged in action relevant processing for a given object domain (Caramazza and Shelton, 1998; Mahon et al., 2009; Mahon and Caramazza, 2011; Konkle and Caramazza, 2017; Conway, 2018). Candidate domains are manipulable (tools) and navigation-relevant objects. On this account, the representations computed in these regions are shape configurations and texture values that statistically reflect the distinguishing properties of manipulable versus navigation-relevant object domains. And, in the measure to which real-world size is differentially correlated with the two object domains (for example, large objects tend to be bulkier) it will have contributed to the evolving preferences for different visual features and configurations in the lateral and ventral occipitotemporal cortex. In other words, the two inanimate-object-preferring regions respond preferentially to, and presumably encode, the types of shape and texture properties that are associated with prototypical manipulable (small, graspable) objects and navigation-relevant (large, stable/stationary) objects.

It is relevant to note here that this view is fully compatible with accounts that emphasize the statistics of visual experience, wherein experienced eccentricity has played an evolutionary role in helping determine the localization of object domain preferences in the brain since it reflects the natural distribution of object viewing: more foveal, required for accurate reaching/grasping for small manipulable objects, versus more peripheral, associated with the requirement for capturing spatial (context) relations for large space-defining objects (Malach et al., 2002, Arcaro et al. 2009; Mahon and Caramazza, 2011; Konkle and Oliva, 2012; Gomez et al., 2017).

A crucial distinction is being drawn here between, on the one hand, the factors that may have determined the observed neural specialization in different regions of OTC for different types of inanimate objects and, on the other hand, the information being computed/represented in those specialized areas. The role of the object property "size" at these two levels of description of neural specialization is illustrative in this regard. It could be argued that real-world object size, because of its role in distinguishing between manipulable and navigation-relevant object types, contributed to help select the visual shape/texture properties that characterize the observed domain specialization in OTC. But what about the role of the visual property "size" in characterizing the computations/representations in these neural regions? Is this visual property explicitly computed/represented in these areas? One reading of the available results is that size is not directly computed/represented in these areas.

There are at least three senses of the property size in the context of visual processing: subtended visual angle, perceived physical size, and real-world size. Konkle and Oliva (2012) found that only objects' real-world size, and not retinal or inferred size, is related to regional specialization. However, this generalization is tempered by the observation, reported in Konkle and Caramazza (2013), that the effect of real-world size is limited to inanimate objects: real-world size differences in the animate domain are not associated with regional specialization. An implication of this observation, consistent with the results reported here, is that it is not real-word size, as such, that drives neural responses but the types of visual shapes that are typically associated with inanimate small, manipulable objects versus large, stable objects (e.g., Long et al., 2016). Furthermore, the property real-world size of an object is a "constructed" value that is retrieved from memory and has no obvious description in a vocabulary of visual properties such a shape, orientation, texture, and color. Indeed, the notion "real-word size" is a kind of conceptual/functional information like the properties "manipulability" and "navigation-relevance", and none of these types of information is directly represented in the putative, size-responsive areas.

By positing a key role for domain-related processing, this view provides an account of data from congenitally blind individuals, who show response preferences in similarly located regions of occipitotemporal cortex when hearing the names of animate versus inanimate objects or big

874 versus small artifacts (e.g., Mahon et al., 2009; Wolbers et al., 2011; He et al., 2013; Peelen et al.,
875 2013; Bi et al., 2015; Mattioni et al., 2020). Since ontogenetic visual experience could not have
876 contributed to the similar patterns of object-preferring effects in high-level visual areas in
877 congenitally blind and sighted individuals, it invites the conclusion that the observed specialization
878 predates such experience. While the domain account provides a principled framework for
879 explaining the potential functional basis for the observed large-scale organization of "visual"
880 cortex, it is silent on the specific representational content at any given level of visual processing.
881 Investigation of the specific perceptual features – e.g., elongated versus bulky shapes, texture
882 properties, curvature patterns, or specific combinations of such features – that are
883 represented/computed in the various object-type-preferring regions is an active, promising area of
884 research (e.g., Andrews et al., 2010; Rajimehr et al., 2011; Bracci and Op de Beeck, 2016; Long,
885 Yu and Konkle, 2018), aided by modeling with deep CNNs (e.g., Güçlü and Van Gerven, 2015;
886 Khaligh-Razavi and Kriegeskorte, 2014; Yamins et al., 2014; Kubilius et al., 2016).

887
888
889
890
891

892 **Acknowledgments:**

893 This work was supported in part by a grant from the Fondazione Cassa di Risparmio di Trento e
894 Rovereto to the CIMeC – Centro Interdipartimentale Mente/Cervello, University of Trento (AC)
895 and by the National Institutes of Health, National Eye Institute (Fellowship F32EY022863-01A1
896 to TK) and was conducted at the Laboratory for Functional Neuroimaging Center at the Center for
897 Mind/Brain Sciences (CIMeC), University of Trento.

898

REFERENCES

Andrews, T. J., Clarke, A., Pell, P., & Hartley, T. (2010). Selectivity for low-level features of objects in the human ventral stream. *Neuroimage*, *49*(1), 703-711.

Ashburner, J., Barnes, G., Chen, C., Daunizeau, J., Flandin, G., Friston, K., ... & Penny, W. (2014). SPM12 manual. *Wellcome Trust Centre for Neuroimaging, London, UK*, 2464.

Auger, S. D., & Maguire, E. A. (2013). Assessing the mechanism of response in the retrosplenial cortex of good and poor navigators. *cortex*, *49*(10), 2904-2913.

Auger, S. D., Zeidman, P., & Maguire, E. A. (2015). A central role for the retrosplenial cortex in de novo environmental learning. *Elife*, *4*, e09031.

Bainbridge, W. A., & Oliva, A. (2015). Interaction envelope: Local spatial representations of objects at all scales in scene-selective regions. *Neuroimage*, *122*, 408-416.

Beauchamp MS (2005) See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. Current Opinion in Neurobiology 15:145–153.

Bi, Y., Han, Z., Zhong, S., Ma, Y., Gong, G., Huang, R., ... & Caramazza, A. (2015). The white matter structural network underlying human tool use and tool understanding. *Journal of Neuroscience*, *35*(17), 6822-6835.

Boronat, C. B., Buxbaum, L. J., Coslett, H. B., Tang, K., Saffran, E. M., Kimberg, D. Y., & Detre, J. A. (2005). Distinctions between manipulation and function knowledge of objects: evidence from functional magnetic resonance imaging. *Cognitive Brain Research*, *23*(2-3), 361-373.

Bracci, S., Cavina-Pratesi, C., Ietswaart, M., Caramazza, A., & Peelen, M. V. (2012). Closely overlapping responses to tools and hands in left lateral occipitotemporal cortex. *Journal of neurophysiology*, *107*(5), 1443-1456.

Bracci, S., & Peelen, M. V. (2013). Body and object effectors: the organization of object representations in high-level visual cortex reflects body–object interactions. *Journal of Neuroscience*, *33*(46), 18247-18258.

Bracci, S., & de Beeck, H. O. (2016). Dissociations and associations between shape and category representations in the two visual pathways. *Journal of Neuroscience*, *36*(2), 432-444.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*(4), 433-436.

Campanella, F., D'Agostini, S., Skrap, M., & Shallice, T. (2010). Naming manipulable objects: anatomy of a category specific effect in left temporal tumours. *Neuropsychologia*, *48*(6), 1583-1597.

930 Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). Region of interest analysis using
931 the MarsBar toolbox for SPM 99. *Neuroimage*, *16*(2), S497.
932 Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The
933 animate-inanimate distinction. *Journal of cognitive neuroscience*, *10*(1), 1-34.
934 Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal
935 cortex for perceiving and knowing about objects. *Nature neuroscience*, *2*(10), 913-919.
936 Chen, Q., Garcea, F. E., & Mahon, B. Z. (2016). The representation of object-directed action and
937 function knowledge in the human brain. *Cerebral Cortex*, *26*(4), 1609-1618.
938 Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is
939 causally and selectively involved in scene perception. *Journal of Neuroscience*, *33*(4), 1331-
940 1336.
941 Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for
942 visual processing of the human body. *Science*, *293*(5539), 2470-2473.
943 Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual
944 environment. *Nature*, *392*(6676), 598-601.
945 Epstein RA (2008) Parahippocampal and retrosplenial contributions to human spatial navigation.
946 Trends in Cognitive Sciences 12:388–396.
947 Epstein, R. A., Bar, M., & Kveraga, K. (2014). Neural systems for visual scene recognition. *Scene*
948 *vision*, 105-134.
949 Gallivan, J. P., McLean, D. A., Valyear, K. F., & Culham, J. C. (2013). Decoding the neural
950 mechanisms of human tool use. *Elife*, *2*, e00425.
951 Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and
952 action. *Trends in neurosciences*, *15*(1), 20-25.
953 Grill-Spector K (2003) The neural basis of object perception. Current Opinion in Neurobiology
954 13:159–166.
955 Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal
956 cortex and its role in categorization. *Nature reviews. Neuroscience*, *15*(8), 536.
957 Güçlü, U., & van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity
958 of neural representations across the ventral stream. *Journal of Neuroscience*, *35*(27), 10005-
959 10014.
960 He, C., Peelen, M. V., Han, Z., Lin, N., Caramazza, A., & Bi, Y. (2013). Selectivity for large

961 nonmanipulable objects in scene-selective visual cortex does not require visual
962 experience. *Neuroimage*, *79*, 1-9.
963 Horn JL (1965) A rationale and test for the number of factors in factor analysis. Psychometrika
964 30:179–185.
965 Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed
966 representation of objects in the human ventral visual pathway. *Proceedings of the National*
967 *Academy of Sciences*, *96*(16), 9379-9384.
968 Julian, J. B., Ryan, J., & Epstein, R. A. (2017). Coding of object size and object category in human
969 visual cortex. *Cerebral Cortex*, *27*(6), 3095-3109.
970 Kalénine, S., & Buxbaum, L. J. (2016). Thematic knowledge, artifact concepts, and the left
971 posterior temporal lobe: Where action and object semantics converge. *Cortex*, *82*, 164-178.
972 Kellenbach, M. L., Brett, M., & Patterson, K. (2003). Actions speak louder than functions: the
973 importance of manipulability and action in tool representation. *Journal of cognitive*
974 *neuroscience*, *15*(1), 30-46.
975 Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models
976 may explain IT cortical representation. *PLoS computational biology*, *10*(11), e1003915.
977 Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and
978 object size. *Journal of Neuroscience*, *33*(25), 10235-10242.
979 Konkle, T., & Caramazza, A. (2017). The large-scale organization of object-responsive cortex is
980 reflected in resting-state network architecture. *Cerebral cortex*, *27*(10), 4933-4945.
981 Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in
982 occipitotemporal cortex. *Neuron*, *74*(6), 1114-112.
983 Kubilius, J., Bracci, S., & de Beeck, H. P. O. (2016). Deep neural networks as a computational
984 model for human shape sensitivity. *PLoS computational biology*, *12*(4), e1004896.
985 Lingnau, A., & Downing, P. E. (2015). The lateral occipitotemporal cortex in action. *Trends in*
986 *Cognitive Sciences*, *19*(5), 268–277. https://doi.org/10.1016/j.tics.2015.03.006
987 Long, B., Konkle, T., Cohen, M. A., & Alvarez, G. A. (2016). Mid-level perceptual features
988 distinguish objects of different real-world sizes. *Journal of Experimental Psychology:*
989 *General*, *145*(1), 95.
990 Long, B., Yu, C. P., & Konkle, T. (2018). Mid-level visual features underlie the high-level
991 categorical organization of the ventral stream. *Proceedings of the National Academy of*

Sciences, *115*(38), E9015-E9024.

Macdonald, S. N., & Culham, J. C. (2015). Do human brain areas involved in visuomotor actions show a preference for real tools over visually similar non-tools?. *Neuropsychologia*, *77*, 35-41.

Mahon, B. Z., Milleville, S. C., Negri, G. A., Rumiati, R. I., Caramazza, A., & Martin, A. (2007). Action-related properties shape object representations in the ventral stream. *Neuron*, *55*(3), 507-520.

Mattioni, S., Rezk, M., Battal, C., Bottini, R., Mendoza, K. E. C., Oosterhof, N. N., & Collignon, O. (2020). Categorical representation from sound and sight in the ventral occipito-temporal cortex of sighted and blind. *Elife*, *9*, e50732.

Mruczek, R. E., von Loga, I. S., & Kastner, S. (2013). The representation of tool and non-tool object information in the human intraparietal sulcus. *Journal of Neurophysiology*, *109*(12), 2883-2896.

Mullally, S. L., & Maguire, E. A. (2011). A new role for the parahippocampal cortex in representing space. *Journal of Neuroscience*, *31*(20), 7441-7449.

Peelen, M. V., Bracci, S., Lu, X., He, C., Caramazza, A., & Bi, Y. (2013). Tool selectivity in left occipitotemporal cortex develops without vision. *Journal of cognitive neuroscience*, *25*(8), 1225-1234.

Perini, F., Caramazza, A., & Peelen, M. V. (2014). Left occipitotemporal cortex contributes to the discrimination of tool-associated hand actions: fMRI and TMS evidence. *Frontiers in human neuroscience*, *8*.

Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling representations of object shape and object category in human visual cortex: The animate–inanimate distinction. *Journal of cognitive neuroscience*, *28*(5), 680-692.

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. (2011). The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biol*, *9*(4), e1000608.

Simmons, W. K., & Martin, A. (2012). Spontaneous resting-state BOLD fluctuations reveal persistent domain-specific neural networks. *Social Cognitive and Affective Neuroscience*, *7*(4), 467-475.

Striem-Amit, E., Vannuscorps, G., & Caramazza, A. (2017). Sensorimotor-independent

development of hands and tools selectivity in the visual cortex. *Proceedings of the National Academy of Sciences*, *114*(18), 4787-4792.

Tarhan, L., & Konkle, T. (2020). Reliability-based voxel selection. *NeuroImage*, *207*, 116350.

Troiani, V., Stigliani, A., Smith, M. E., & Epstein, R. A. (2014). Multiple object properties drive scene-selective regions. *Cerebral cortex*, *24*(4), 883-897.

Wang, X., Peelen, M. V., Han, Z., He, C., Caramazza, A., & Bi, Y. (2015). How visual is the visual cortex? Comparing connectional and functional fingerprints between congenitally blind and sighted individuals. *Journal of Neuroscience*, *35*(36), 12545-12559.

Wickham, H (2009). ggplot2: elegant graphics for data analysis. *Springer New York*, *1*(2), 3.

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619-8624.